

A Little Give and Take

Adam Oliver

London School of Economics and Political Science

a.j.oliver@lse.ac.uk

Abstract

In this article, I contend that the behavioural effects that tend to be labelled as errors by most behavioural economists, and as such have served as the justification for a paternalistic direction in behavioural public policy (i.e. policy intervention that aims to protect people from imposing harms on themselves), are in an ecological sense not errors at all. While acknowledging that modern societies are very different from the types of societies in which these effects evolved, I argue that we still cannot conclude that attempts to modify people's choices in accordance with these so-called errors will improve the lives of those targeted for behaviour change, particularly given the varied and multifarious private objectives and desires that people pursue. Where people are imposing no substantive harms on others, I maintain that policy makers should restrict themselves to protecting and fostering the fundamental motivational force of reciprocity, which serves to benefit the group (which could be the whole society) and, by extension, most of the people who comprise the group, irrespective of their own personal desires in life. However, when one party to any particular exchange actively uses the behavioural affects to benefit themselves but imposes harms on the other party to the exchange, the concept of a free and fair reciprocal exchange has been violated. In these circumstances, there is an intellectual justification to introduce behavioural-informed regulations – a form of negative reciprocity – against activities that impose unacceptable harms on others. My arguments thus call for behavioural public policy to preserve individual autonomy within an overarching policy framework that nurtures reciprocity whilst at the same time regulates against behavioural-informed practices that impose substantive harms on others, rather than focusing on reducing the harms that people supposedly impose on themselves. This would be a major switch in emphasis for one of the most important developments in public policy in modern times.

Keywords: Behavioural, Ecology, Evolution, Prospect theory, Reciprocity, Reflection, Scarcity

Introduction

The motivational force of reciprocity – of responding in kind to good and bad intentions and/or actions – can and should be nurtured by policy makers to help individuals pursue their own personal desires and to help public sector groupings achieve their collective objectives [1]. The urge to act reciprocally lies deep within the human psyche, and evolved because the long-term self-interest of individuals is often furthered by acting in a way that brings forth benefits and protections to the group as a whole. Out of this evolutionary process arose a social norm that favours conditional cooperation and that justifies the punishment of transgressors.¹

Admittedly, that the members of subgroups within society also often act cooperatively and reciprocally among themselves but to the detriment of their wider group or society at large is an unfortunate, and long recognised, possibility. Hume, for instance, wrote that ‘Robbers and pirates ... could not maintain their pernicious confederacy, did they not establish a new distributive justice among themselves, and recall those laws of equity, which they have violated with the rest of mankind’ [2]. There are also risks associated with negative reciprocity, including undue or excessive retribution and spiralling retaliation, but, if harnessed in the right way, positive and negative reciprocity can both serve substantively as a force for good.

My intention in this article is to contend that the motivational force of reciprocity, if protected and nurtured by policy makers, can serve to benefit individuals, irrespective of what their personal desires in life may be. This, I will argue, is a more appropriate policy approach than that which currently dominates the burgeoning field of behavioural public policy (which is, incidentally, one of the major developments in public policy in modern times): namely, assuming that people often make decisions that run counter to their own best interests, so legitimising policy makers to force or guide them in alternative directions. The protection of individual autonomy is thus central to my thesis.

However, I will also argue that some people will inevitably try to take advantage of others in circumstances that place a high premium on autonomy, to the extent that the concept of a free and fair reciprocal interaction between exchange partners will sometimes be violated. In these circumstances, behavioural-informed regulations against harms – a form of negative reciprocity – might be warranted. But first, since reciprocity plays a central role in my arguments, I offer some further reflections on the origins of this motivational force.

The Origin of Reciprocity

Those who write on reciprocity present varied, if interrelated, explanations for its origin. Henrich, for example, wrote that reciprocity underpins the mutual protection that became necessary after our ancestors descended trees and became ground apes, and Boehm noted that reciprocal tendencies strengthen when individual success in a hunt is uncertain [3,4]. He

¹ For further discussion of evolutionary mechanisms appertaining to reciprocity, see Michael Muthukrishna’s article in this issue.

reported that when chimpanzees hunt, those that gain initial control of the carcass will share just enough to enable them to retain control, and that there may be reciprocation between givers and receivers when their relative success is reversed in the future. According to Boehm, compared to chimpanzees, early homo sapiens killed larger game and thus there was more sharing, while in hunter-gatherer societies dominance over meat was often negated entirely by it being shared out by a neutral person.²

In political philosophy, there is a long history of debate on the cause, purpose, extent and form of reciprocity. That reciprocity evolved to serve mutual self-interest was, for instance, an important feature of Hume's suggestion that the cooperation that is necessary to the subsistence of small family units extends eventually to encompass whole societies [2]. However, it is important to emphasise that those who attach much importance to reciprocity as a motivational force do not deny that humans are influenced by a mix of motives, including more selfishly-driven short-term egoism as well as perhaps even pure altruism, or that the relative strengths of these motivations vary both interpersonally and – over context and time – intrapersonally. It thus ought to be recognised that there is the potential for the selfishly egoistic driver, if one is not careful, to crowd out the notion of give and take, which could ultimately be detrimental to the group or wider society – and by extension to the individuals of whom the group is comprised.³

In my own field of expertise, behavioural economists tend not to take an evolutionary approach to human decision-making; rather, most behavioural economists accept the set of assumptions that are postulated by standard notions of rational choice as normatively valid, and then endeavour to demonstrate that in their actual choices, people often systematically violate those assumptions. Given this general acceptance that rational choice theory assumptions are normatively correct, such violations are treated as errors in individual decision-making, and this, in turn, provides the justification for most of those working in the burgeoning field of behavioural public policy to adopt a paternalistic framework: i.e. that since people are error-prone in their decision-making, they need to be helped to correct those errors. However, if one were a little more open to evolutionary explanations for the heuristics that people follow when making their decisions, one might realise that there may be good reasons for why the phenomena that apparently cause these 'errors' emerged, and moreover, that these reasons may remain valid in modern contexts. In short, evolutionary explanations may put into question the validity of normative assumptions held by most behavioural economists.

For example, most working within the field of behavioural public policy attribute present bias – the emphasis, oft perceived as unreasonable, that people place on the immediate moment –

² Wiessner, although acknowledging that her focus is on just one hunter-gatherer society – the Ju/'hoansi of the Kalahari – reported that successful hunters do preferentially give meat to others, particularly towards those to whom one is obligated and to close kin so as to create pleasurable living arrangements, and against those who are unwanted by the society (which respectively hint at positive and negative reciprocity) [5].

³ Although we might conclude that in hunter-gatherer societies (and before), reciprocity evolved for the good of the group and its individual members, as societies grew and became more atomised, more opportunities arose for the egoistically inclined to act upon their motives with less fear of being detected. Thus, a form of social contract that incorporates threats of negative reciprocity – that is manifested in most of the world's major religious codes and is embedded in laws – became necessary to reinforce the socially beneficial norm of positive reciprocity.

to a failure of willpower. However, this tendency perhaps evolved as an appropriate response to circumstances where one's future was typically uncertain and possibly bleak [6]. In short, if one does not perceive much of a future, why not focus upon the present? Of course, while this may have been a reasonable strategy over much of our evolutionary history, in modern, relatively secure societies, present bias can cause a range of actions and inactions that are evidently harmful to our future selves, from smoking too much tobacco, to consuming too much alcohol, to saving too little for retirement. However, it is something of a leap of faith to believe that people, even in modern contexts, are acting irrationally if they do not place as much weight on the future as most behavioural public policy analysts might wish them to.

If we are not willing to take this leap of faith, it is because we cannot conclude that third parties are in a good position to determine that the behavioural heuristics and affects are necessarily causing errors in an individual's decision-making, with respect to the individual's own desires in life. As aforementioned, saving insufficiently for retirement is often attributed to present bias, and this serves as a justification for introducing opt-out pension plans. However, such pension plans may be detrimental to a great many people for various reasons. For instance, some people may simply prefer not to be enrolled, yet may not realise they have been manipulated into doing so, while an opt-out plan may disincentivise others from searching around for plans that might better suit their desires. Of course, it may be the case that the behavioural affects that evolved in different circumstances might sometimes lead people living in modern contexts astray, but I argue that it is better to try to educate people about these affects and then leave them to their decisions than to attempt to use manipulative or coercive paternalistic measures to move them in particular directions "for their own good", when we cannot really be sure that these directions really serve their own good at all.

Furthermore, in many experiments, behavioural patterns perceived broadly as irrational have focussed on individual decision-making, yet in the real world, collective responses to bolster individual security have evolved. Behavioural experiments are often ill-equipped to uncover these collective responses, but even with respect to those experimental designs that are open to other-regarding and reciprocity-motivated actions (such as ultimatum games, trust games, public goods games etc.), such actions often tend to be viewed as irrational because they conflict with the dominant model of selfish egoism.⁴

Behavioural economists typically argue that prospect theory holds humans to be systematically error-prone, particularly with respect to responses that are driven by risk attitudes. However, it may be that within the highly stylised framing of typical prospect theory tests, behavioural patterns consistent with that theory's predictions are driven by deeply ingrained survival instincts that evolved in relation to perceptions of scarcity and abundance of food resources.⁵ I will elaborate on these points below.

⁴ The economic history on what ought to and does fundamentally motivate human beings is nuanced and complex [7]; however, most behavioural economists and behavioural paternalists appear to accept that economic rationality equates to individual egoism.

⁵ Nutritional needs are the most basic requisites of life. Everything else, including reproductive success, depends on their being met. If you are faced with two or more options and only one of those options offers the opportunity for you to meet your survival needs, then that option is the one that you will have to choose – in those situations, relative risk considerations are irrelevant.

On Reflection

Prospect theory makes two main modifications to the standard theory of rational choice: First, rather than final assets, the subjective carriers of value are assumed to be gains and losses around a reference point (with the reference point generally perceived to be the status quo, the most likely or expected outcome, or the aspiration level),⁶ and with losses weighted roughly twice as much as gains of the same magnitude (implying loss aversion); and second, it assumes the subjective weighting of probabilities, such that low probabilities are overweighted and high probabilities are underweighted, rather than processed in their objective mathematical form.

Both of these insights have been used prominently in policy design. For instance, in terms of offering people small incentives to change their behaviours with respect to smoking, exercise, or diet, deposit contracts – which require individuals to deposit small amounts of money that is only returned to them if they meet their behaviour change targets – and lottery incentive mechanisms – which offer a small chance of winning a sizeable amount of money if people meet their behaviour change targets – have both been quite extensively trialled, and their designs are respectively informed by loss aversion and probability weighting.

The creators of prospect theory, Tversky and Kahneman, stated that the ‘most distinctive implication of prospect theory is the fourfold pattern of risk attitudes’ known as the reflection effect, which is summarised in Table 1 [8].

Table 1: The Reflection Effect

	Gains	Losses
High Probability	Risk Aversion (Fear of Missing Gain)	Risk Seeking (Hope to Avoid Loss)
Low Probability	Risk Seeking (Hope of Gain)	Risk Aversion (Fear of Loss)

The top left quadrant in Table 1 describes the prospect theory risk attitude prediction when a person is faced with a large probability of a gain – for example, a 0.99 chance of winning £100. If an individual is offered a choice between this risky option and the certainty of its expected value of £99 (i.e. 0.99×100), prospect theory predicts that the individual will place a high weight on the certainty (in part because the subjective value of gaining £100 will only be very marginally greater than the subjective value of gaining £99, and in part because people will overweight the 1% chance of winning nothing), will reject the risky option and will thus be risk averse. The rules of rational choice theory assume that people will be close to indifferent between a gamble and the certainty of its expected value, but these rules implicitly assume that people are able to repeat the game many times. However, in evolutionary terms

⁶ For instance, in standard rational choice theory, the satisfaction that a man feels for earning a bonus of £10,000 is assumed to be independent of his aspirations, but prospect theory allows for him to feel differently towards the £10,000 if he aspired to a bonus of, say, £5,000 compared to an aspiration of £15,000.

(e.g. securing game in a hunt for sustenance over the next few days), and even today, many decisions may be “one-shot” (or at most, “limited-shot”). In such circumstances, one may quite rationally choose to avoid a risky option even if the gamble is heavily weighted in one’s favour if sufficient resources can be secured without risk, and conversely, one might engage in an unpromising risk if the alternative is certain doom. These issues will be further discussed later.

The bottom left quadrant describes the predicted risk attitude when a person is faced with a small probability of a gain, such as a 0.01 chance of £100. Here, prospect theory predicts that the individual will overweight the chance of winning, would hence prefer the gamble over its expected value of £1 (i.e. 0.01×100), and will therefore be risk seeking. The top and bottom right quadrants can be read similarly, and show that prospect theory predicts opposing risk attitudes for losses as compared to gains for both large and small probability scenarios.⁷ That the predicted risk attitudes are reflected across gains and losses for both high and low probabilities gives the reflection effect its name, and this fourfold pattern of risk attitudes contrasts with that of universal risk aversion or risk neutrality predicted by standard rational choice theory.

By asking respondents to choose between gambles and their expected values, Tversky and Kahneman reported some empirical support for the full fourfold pattern of risk attitudes, but, perhaps strangely given the importance of the reflection effect to modern behavioural economic theory, controlled testing of the full effect is quite scarce and, where it does exist, is mixed (see [9] and the references therein). One possible reason why the evidence is mixed is that prospect theory, in assuming that people will assess a risky option by mechanistically weighting the subjective value of its outcomes with their associated subjective probabilities, may not take into account the psychological processes that have evolved to help humans to deal with uncertainty in the face of perceptions of scarcity and abundance.

Anatomically modern humans emerged 200,000 years ago and hunter-gathering was the dominant form of social organisation until the development of agriculture about 12,000 years ago. The consideration of well-defined probabilities (whether or not people subjectively weight them), as used in games of chance, stretches back 3,000 years or so, although through the avenues of formal education and pastimes such as gambling on sporting outcomes, the widespread exposure of the members of society at large to well-defined probabilities is more recent. It is thus not implausible that the way in which humans deal with uncertain events now is still influenced heavily by the processes that evolved to help our hunter-gatherer ancestors (and their predecessors) deal with uncertainty in their search for food, which are likely to have been driven by perceptions of frequency of success based on their recent experiences, and on the magnitude of outcome.

For instance, if the likelihood of securing a good catch was high but posed a non-negligible element of risk, such a pursuit may implicitly have a high expected value; however, there would still be a chance of failure, which could have catastrophic consequences. An alternative strategy that promised a less impressive but still sufficient catch but with less or

⁷ For example, in relation to the top right-hand quadrant, if an individual were offered a choice between a 0.99 chance of losing £100 or losing its expected value of £99 for sure, prospect theory predicts that the risky option will be chosen (i.e. risk seeking behaviour); with respect to the bottom right-hand quadrant, prospect theory predicts that a choice between a 0.01 chance of losing £100 and losing its expected value of £1 for sure would result in a preference for the sure loss (i.e. risk averse behaviour).

no risk attached to it may well have had a lower expected value (in terms of caloric content), but if it had a higher chance of sustaining life by guaranteeing sustenance, it would be a perfectly sensible strategy to pursue, particularly in the absence of any facility for the long-term storage of meat. The focus of hunter-gatherers may well have been on securing sufficient food to sustain themselves over the relative short-term, rather than the maximisation of expected value. Conversely, if food was thought to be scarce due to infrequent recent hunting successes and/or modest prey, then people may have realised that unless they took risks there would be an insufficient catch to sustain themselves. In short, through necessity they would have been driven to seek risk. The hypothesis of implicit risk aversion in the face of abundance and implicit risk seeking in the face of scarcity to facilitate survival is the same as that postulated by risk sensitivity theory, which behavioural ecologists use to explain rational food acquisition decision-making by foraging animals in environments that are highly uncertain [10].

The two scenarios just described – i.e. high and low frequencies of success, suggesting relative abundance and relative scarcity – respectively mirror the top and bottom left quadrants of Table 1 (i.e. aspects of the reflection effect discussed earlier), where probability serves as a proxy for prior frequency of success. If people similarly perceive the top and bottom right quadrants of Table 1 as scenarios that respectively indicate relative scarcity and relative abundance, then we may have an explanation for the behaviour patterns that are consistent with the full reflection effect. For instance, a high probability loss – the top right quadrant – might invoke perceptions of relative scarcity, which would provoke behaviours that are consistent with risk seeking. In contrast, a low probability loss – the bottom right quadrant – might imply relative abundance if a loss is interpreted as anything less than one's aspiration levels (or a return that meets the forager's energy requirements, in the framing of risk sensitivity theory). In this latter scenario, behaviours that are consistent with risk seeking would not be expected, because to facilitate survival (which is, in an evolutionary sense, one's primary objective) they are not necessary (and indeed may well be harmful).⁸

Behaviours that are consistent with the predictions of prospect theory are therefore not necessarily as erroneous as is often assumed; they are perfectly rational when one adopts an ecological perspective. That is not to say that risk seeking in the face of scarcity will guarantee survival, but that it may in some circumstances be the only strategy that offers any chance of survival. To further improve one's chance of survival, however, collective strategies evolved over the course of time – strategies that the individualistic experimental frameworks used in much of the behavioural economic literature, including those employed in tests of the reflection effect, are ill suited to uncover.

Signalling Cooperative Intent

⁸ Mishra wrote that '... the classic finding [of prospect theory] that decision-makers are risk-averse in the face of gains and risk-prone in the face of losses in framed decision scenarios may be a by-product of decision-makers seeking to minimise the possibility of experiencing a negative outcome that does not meet their needs' [10]. The prospect theory predictions are actually somewhat more nuanced than those suggested by Mishra in this quote, but the same basic point applies with respect to the four-way pattern of risk attitudes predicted by the reflection effect. That said, the size of the outcome might sometimes be the dominant factor in influencing perceptions of scarcity and abundance. For example, if there is a good chance of experiencing what might be considered a gain but the gain is tiny, the perception might be one of relative scarcity rather than relative abundance, which may thus provoke actions and behaviours that are consistent with risk seeking rather than risk aversion. Thus, it is possible that the consideration of either extremely small or extremely large outcomes might lead to ecologically rational choices that are inconsistent with the predictions of prospect theory.

Fortunately, experimental games have been developed that suggest that prosocial, collective motivations are embedded deep in the human psyche, which in turn offer food for thought with respect to policy design (see [1]). Such cooperative intent is unfortunately also often misattributed as a failure of rationality in the behavioural economics discourse. As noted earlier, these experiments include the ultimatum, trust, and public goods games, but for illustrative purposes let us here consider the so-called centipede game in a little detail.

Consider two players, A and B, and four stages, I to IV. In stage I, A chooses to ‘take’ or ‘pass’ – if A takes, he receives 80% of an initial endowment of, say, £5 (and player B receives the remaining 20%). If he passes, the endowment doubles in size (to £10) in Stage II, and B gets to choose whether to take or pass. Likewise, if B takes, she gets 80% of the growing endowment, leaving 20% to A, but can pass back to A if she wishes. If she passes, the endowment is again doubled in Stage III, with A left with the choice of taking or passing. Assuming neither has taken, the stages are completed at Stage IV. Table 2 summarises the above.

Table 2: The centipede game

	Takes	Passes
I (A decides)	(£4, £1)	II
II (B decides)	(£2, £8)	III
III (A decides)	(£16, £4)	IV
IV (B decided)	(£8, £32)	Finish
Finish	(£64, £16)	

The table shows that if A takes in stage I, he will get £4 and leave £1 to B, but if he passes then B, in stage II, can either take £8 and leave £2 to A, or pass, which would place the ball once again in A’s court, etc. If the players were to proceed to the finish, A and B would respectively receive £64 and £16, leaving both players considerably better off than if the game had ended at stage I. However, B would be better off taking rather than passing at stage IV (as she would receive £32 rather than £16), and thus (behavioural) economists typically conclude that rationality requires her to take at that point. Since A would be better off taking in stage III than if the game was to end with B taking in stage IV, economists tend to further assume that A would take if he got as far as stage III; but a “rational” B would know this and she would thus be better off taking in stage II. Through backward induction, a fully economically rational person would always decide to take. Rational choice theory therefore predicts that the game will end with A taking in stage I. However, McKelvey and Thomas undertook an experiment where 93% of players passed in stage I, 62% passed in stage II, 35% passed in stage III, and 25% passed in stage IV [11]. It appears that in the centipede game, the players will cooperate, at least up to a point, to try to ensure that they will each benefit more than would otherwise be the case. Signalling and acting upon deep cooperative instincts that evolved due to their propensity to serve mutual self-interest might explain the decisions that are often erroneously labelled as irrational in this type of task.⁹

⁹ It can rightly be contended that cooperation substantially diminished through the stages of the centipede game in the results presented by McKelvey and Thomas, but note that the centipede game includes a clear, defined final stage that incentivises selfish self-interest as the end approaches. In the real world, collective reciprocal exchange between individuals often has no clearly defined endpoint, and thus the incentives for continued cooperation are less likely to erode so readily.

If the interpretations offered in this article are correct, then when facing a situation of scarcity individual risk averse strategies may not secure sufficient resources for survival and thus individual risk seeking strategies are necessary. Of course, total reliance on such strategies are still likely to result in insufficiency, and, consequently, collective strategies evolved in those instances where scarcity was not so extreme as to render sharing a threat to immediate survival. Such strategies meant that those who were fortunate to obtain sustenance at a particular moment shared in the expectation of reciprocation when the fortunes of the parties were reversed. In short, individual risk seeking is borne out of necessity (cf. reflection effect experiments) and collective risk aversion – i.e. reciprocal sharing, a form of rudimentary insurance – evolved to enable us to experience a more constant level of security and to mitigate the misfortune to which we might be exposed if we relied entirely on ourselves (cf. cooperative game experiments).¹⁰

With the evolution of reciprocal cooperation, there of course remained incentives for people to act in a more selfishly egoist manner if they thought they could get away with it. Thus, negative reciprocity – the threat and act of punishment – at least in part emerged to crowd in the broadly beneficial acts of positive reciprocity among those who might otherwise transgress. In short, to protect security in general, some specific freedoms had to be constrained.

Constraining Freedom

I have tried to suggest in this article that the supposed behavioural anomalies that influence individual decision-making, such as present bias – which are often deemed irrational by behavioural paternalists – probably evolved for legitimate reasons. The modern world is of course quite different from the circumstances in which the behavioural affects evolved. For instance, people living in hunter-gatherer societies did not have to worry about saving for their retirements decades into the future, but even in the modern context, a third party (e.g. a policy maker) cannot really know whether the citizenry would be better off by saving, rather than spending, in the present moment, as judged against each citizen's own objectives and desires in life. It is therefore the contention here that manipulating or coercing people into pension plans is an inappropriate course of action. Rather, attempts ought to be made at better educating people with respect to the implications of saving and not saving, while leaving them to make their own decisions in relation to these (and other) matters.

That said, I contend here that an evolved sense of pro-social, reciprocity-driven self-interest has emerged that ultimately benefits most people in society, irrespective of their own personal objectives and desires in life. A legitimate role for policy makers is therefore to shape society in broad terms, and our institutions and their policies specifically, so as to encourage, and to certainly not discourage, these reciprocal instincts. As an example, modern pension systems are essentially informed by the principle of intergenerational indirect reciprocity (even if most people might not be aware of this). That is, people of working age pay into the system now so that those who are already retired receive their pensions; when they themselves have retired, future working populations will similarly finance their pensions. Making people more aware of the reciprocal underpinnings of pension systems –

¹⁰ This explanation for the evolution of reciprocal sharing is the same as that offered by Boehm, presented earlier [4].

i.e. you pay forward to others so that others pay forward to you – might bode well for the sustainability of pension systems.¹¹ So long as the general structures of society are conducive to cooperative behaviours, there is, I suggest, no call for policies that interfere too much in the choices that people make (so long as those choices are not imposing substantive harms on others); if you allow people to be free, most have the evolved mental apparatus to seek and find practices of mutual benefit without the hands-on, top-down involvement of third parties who claim to know better.

Although a third party cannot really discern if and where the behavioural affects cause people to impose net harms upon themselves in their passive decision-making, it is possible to identify fairly common circumstances where particular actors, out of self-interest, actively take advantage of these affects by interfering in the notion of a free and fair (reciprocal) exchange. By doing so, they impose harms on their exchange partners. For example, less than reputable financial institutions encourage people to take out short-term high-interest loans by emphasising the joys of spending in the moment and concealing the long-term pain of repayment, and online gambling companies use a gamut of behavioural affects, including anchoring, probability weighting, manipulations around the perception of losses, and prestige effects by employing (millionaire) celebrities in their advertising campaigns, to entice people to start and continue activities that could soon place them in very serious financial difficulties.

Countless other examples of one party to an exchange using the behavioural affects to manipulate the other party could be introduced – indeed, many instances can be identified if one limits oneself to the food and drinks industry alone. For example, the marketing divisions of confectionary companies know from long experience that salience and immediacy can have a substantial impact on consumer buying patterns, and have traditionally paid supermarkets to display their products at child eye-level close to checkout counters. Similarly, supermarkets are often financially induced to place alcoholic beverages at the end of shopping aisles since it is known that this salient positioning of products substantially increases sales. Finally, in a recent blogpost, the money savings expert Martin Lewis observed that his local supermarket was selling single packs of Jaffa Cakes (each containing ten cakes) next to double packs (containing twenty cakes), with both pack sizes priced at £1.¹² This is an example of what behavioural economists have termed the decoy effect, whereby retailers offer a choice of items where one of those items is, relatively speaking, obviously a bad deal, in the expectation that the other item will now appear almost too attractive to resist. The introduction of the decoy thus potentially impacts negatively on a free and fair exchange – a person might not really want any Jaffa Cakes, but with the decoy in place might be manipulated into buying twenty of them.

Such acts of manipulation are borne out of egoism. One reason to curtail freedom is to prevent or punish those who might, often implicitly, use the instruments of behavioural science to benefit themselves through imposing unreasonable harms on others. Therefore, although behavioural paternalism is not an evidently appropriate strategy for improving people's lives given the varied and multifarious objectives and desires that each of us may pursue (e.g. if a person purchases a pension plan in a free and fair exchange then there is no reason to prevent or covertly guide him from doing so) there is a legitimate role for

¹¹ In a recent issue of this journal, I offered some further suggestions for how governments might nurture the reciprocal instincts (see [7]).

¹² At the time of writing, the blogpost can be viewed at: <https://blog.moneysavingexpert.com/2020/09/martin-lewis--i-confess--i-did-a--decoy-effect--experiment-on-my/>

government regulation against practices that use the behavioural affects for pernicious purposes (e.g. to manipulate a person into purchasing a particular pension plan).

In sum, I will conclude this article by reiterating what I see as the most appropriate way forward for the burgeoning field of behavioural public policy: specifically, a direction that preserves individual autonomy within a policy framework that nurtures reciprocity, whilst at the same time regulates against behavioural-informed practices that impose unreasonable harms on others.

Acknowledgments

I thank Michael Muthukrishna and Jennifer Sheehy-Skeffington for guiding me in the direction of literature that is relevant to my arguments, and Irene Bucelli and Nicholas Reed Langen for editorial comments.

References

- [1] Oliver, A. 2019. *Reciprocity and the Art of Behavioural Public Policy*. Cambridge: Cambridge University Press.
- [2] Hume, D. 1751. An Enquiry Concerning the Principles of Morals. In: Coventry, A., and Valls, A. (eds.). 2018. *David Hume on Morals, Politics, and Society*. New Haven: Yale University Press.
- [3] Henrich, J. 2016. *The Secret of our Success: How Culture is Driving Human Evolution, Domesticating Our Species and Making us Smarter*. Princeton: Princeton University Press.
- [4] Boehm, C. 2012. *Moral Origins. The Evolution of Virtue, Altruism and Shame*. New York: Basic Books.
- [5] Wiessner, P. 2002. Hunting, Healing and Hxaro Exchange: A Long-Term Perspective on !Kung (Ju/hoansi) Large-Game Hunting. *Evolution and Human Behavior* 23: 407-436.
- [6] Pepper, G. and Nettle, D. 2017. The Behavioural Constellation of Deprivation: Causes and Consequences. *Behavioral and Brain Sciences* 40: e314.
- [7] Oliver, A. 2021. An Unhappy Pursuit of Happiness. *LSE Public Policy Review* 1(3): 3.
- [8] Tversky, A., and Kahneman, D. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* 5: 297-323.
- [9] Oliver, A. 2018. Your Money and Your Life: Risk Attitudes Over Gains and Losses. *Journal of Risk and Uncertainty* 57: 29-50.
- [10] Mishra, S. 2014. Decision-Making Under Risk: Integrating Perspectives From Biology, Economics and Psychology. *Personality and Social Psychology Review* 18: 280-307.
- [11] McKelvey, R., and Palfrey, T. 1992. An Experimental Study of the Centipede Game. *Econometrica* 60: 803-836.